

## Listening to the Ensemble of Story-Telling Performance and Audience Reactions

Ryota Nomura

The University of Tokyo  
E-mail: [nomuraryota@gmail.com](mailto:nomuraryota@gmail.com)

Doi:10.5901/ajis.2013.v2n9p597

### Abstract

*In vaudeville settings, story-telling performers and their audiences mutually depend on each other. While the audience often smiles in response to incongruent lines and interpreted gestures, the performer sometimes delivers the punch lines only after audience-initiated smiles and movements. However, these temporal structures are too complex to capture directly. In this study, I have proposed a rigid method of formulation and a method of listening to the implicitly organized temporal patterns between the performance and audience reactions. A professional Rakugo storyteller performed live in a laboratory in front of 20 audience members aged 16 to 67 ( $M = 40.6$ ,  $SD = 16.4$ ). The qualitative observation of motions and voices were converted to quantitative intensities by using computer programs. I also detected where each audience member concentrated his or her attention, by focusing on eyeblinks, and assigned the identical musical pitch to the reactions of the same person. As a result of mixing these data sequentially along with the time series, an actual musical score was composed. This music provided rich data in which we could easily recognize coordination patterns in the vaudeville setting as a temporal gestalt. The synchronization of reactions among audience members was represented as forming a chord and time-delayed coordination between performer and audience, which sounded like interchanges. It is suggested that the musical perspective introduced in this study could lead to the development of assessment techniques for a variety of orators that regularly speak in front of audiences.*

**Keywords:** performing arts; appreciation; performer-audience system; temporal gestalt; eyeblink

### 1. Introduction

In vaudeville settings, story-telling performers and their audiences are mutually interdependent. While the audience often smiles in response to incongruent lines and interpreted gestures, the performer sometimes delivers the punch lines only after audience-initiated smiles and movements. The interactive nature between the performer and the audience forms the basis for the collective emotional experience of the performance. The enjoyment of a live vaudeville performance is not just the fun of listening to jokes. Rather, it is the whole body immersive experience involving audio-visual, cutaneous, and proprioceptive sensations (Nomura & Maruno, 2011). When a performer and the audience share a spatial and temporal field, they collaboratively create one performance. In such situations, they share joint responsibility for the performance through rich interactions that are based on multimodal communication. Although all story-telling performers must know about the interactive nature of their performance at one level or another, there is very few empirical research on this issue, perhaps because of the intrinsic difficulties in conducting such studies.

Developing a method of capturing the interaction between performers and the audience has been an enduring and fundamental issue in research on the performing arts. There are no visible links in actual interactions; therefore, we must recognize phase changes in interactions through representations. Additionally, even if we were successful in making such interactions visible, a critical problem would still remain. It would be difficult for a person to visually recognize the complex temporal gestalt, because of the realistic limitations on cognition. When a picture is drawn to represent a temporal gestalt in interaction, the links among the elements of the picture increase exponentially with the increasing number of interactions, leading to visual representations that become increasingly dense. Therefore, we may not be able to visually capture the phase and its changes along with the time series.

The fundamental and unique idea presented in the current study was capturing phases of an interaction as music. Regardless of the complexity of an interaction, it can be expressed as a superposition of sound waves. Humans can comprehend such acoustic expressions much easier than visual representations and are able to identify such phases as a temporal gestalt. The high-quality resolution of sounds is implemented by the beautifully elaborated inner ear system (Corey et al., 2004). Therefore, I have proposed a method for capturing interactions in a story-telling performance as a musical expression. The method focused on both the performance and audience reactions. It utilized a performer's

movements and quantified audience reactions, based on their eyeblinks when they were watching a performance. As I have discussed below, eyeblinks of the audience reflect how members of the audience perceive and interpret a performance. The purpose of the current study was to propose a method that would make the performer-audience system "listenable," and to test the validity of this method from the perspective of its logical and analytical value for research.

## 2. Previous research and the current study

### 2.1 *Appreciation of performing arts in vaudeville settings*

The inherent and core nature of the performing arts, including story telling, is that the performance is temporally continuous. In the process of appreciating a performance, the audience has to make some particular segments on the perceived performance according to each person's way of viewing, in order to understand the meaning and give an interpretation. The place where each member of an audience input the segments is arbitrary. However, segmentation by the audience would be coherent among the audience members, if the reference framework for the segmentation is common. Some expert story-telling performers can simultaneously draw many members of the audience into the world of the story, and thereby induce similar emotional responses from the audience. Especially in vaudeville settings, experts prepare clues, not only for understanding the story, but also for smiling and laughing. Experts can then communicate their intention to increase smiling and laughter without directly mentioning their intention. In other words, the performer and audience interact with each other at a meta-communication level (Sanches, 1975). If the audience does not receive the unmentioned message, the performers would modify their own face, voice, or motion in concordance with the audience sensitivity. In this sense, a story-telling performance is not a unidirectional transfer from the performer to the audience, but it is jointly constructed through interactions between the performer and the audience. The audience is not aware that they are attending a bi-directional communication, however, they actually cooperate with it by applauding and whistling intentionally, and by laughing and swaying their bodies unintentionally. Moreover, audience members frequently blink their eyes as an essential response to the performance. Empirical studies on the performing arts have not investigated the role of eyeblinks. Nevertheless, eyeblinks are important in vaudeville settings. It has been pointed out that a decrease (lower frequency) of eyeblinks is suggestive of the allocation of attention (Stern et al., 1984). Therefore, researchers can easily and objectively identify when and how the attention of an audience is controlled by the action of the performer.

### 2.2 *Eyeblink entrainment and speaker-listener coordination*

Eyeblinks are usually non-intentional and voluntary, periodic movements. Whereas the periodicity of eyeblinks is driven by internal physiological conditions of each individual, it is altered by external stimuli (Stern et al., 1984). We usually generate spontaneous eyeblinks every few seconds, much more often than necessary for ocular lubrication. Some studies have pointed out that the timings of eyeblinks are related to attentional breaks (Nakano, Kato, Morito, Itoi, & Kitazawa, 2013). For example, when people read a sentence, their eyeblinks increase at the punctuation marks (Hall, 1945). Participants that try to accomplish difficult tasks, also increase their eyeblinks (Stern et al., 1984). In recent years, Nakano and her colleagues have reported a series of new findings on spontaneous eyeblinks in communication (Nakano, Kato, & Kitazawa, 2011; Nakano, Yamamoto, Kitajo, Takahashi, & Kitazawa, 2009; Nakano, & Kitazawa, 2010).

In one experiment, the eyeblinks of participants synchronized when they watched a comedy film. Participants blinked frequently when the main character was absent, when watching scenes in long shot and during repeated presentations of a similar scene (Nakano et al., 2009). Such synchronization was not found when participants watched a background video, or when they listened to a story read aloud (Nakano et al., 2009). Also, it has been suggested that the synchronization of eyeblinks is related to mutual coordination during face-to-face communication. Nakano and Kitazawa (2010) demonstrated "eyeblink entrainment" in situations in which the listeners viewed close-up video clips (with sound) of the speaker's face. They reported that listeners blinked after a 0.25 to 0.5-second delay after the speaker's blink. In a following study, Nakano et al. (2011) reported that eyeblink entrainment was not observed in individuals with autism spectrum disorders, who are characterized by impaired social communication. According to Nakano et al. (2011), these results suggest that individuals with ASD are unable to temporally attune themselves to the behaviors of others. In sum, temporal coordination may facilitate effective social communication with others. Therefore, eyeblinks could be an appropriate index for coordinating face-to-face social communication. In this study, I have proposed a method for quantifying interactions resulting from the temporal coordination between the speaker (performer) and the listener (audience) by focusing on the eyeblinks of audience members.

### 2.3 Design of an ensemble: Making a temporal gestalt listenable

Many audience members would be watching a performance in a vaudeville setting; therefore, a method was devised to identify instances in the performance, when each audience member blinks simultaneously with others, and to adequately express this on a time sequence. In the method, I assigned an arbitrary musical tone to every audience member, such that at each eyeblink, when the audience members close/open the eyelids, the assigned tone would be either on, or off (i.e., a period of blink stroke one sound) along the time course of the performance. Thus, eyeblinks of several audience members would form an ensemble, starting to collectively play a non-intentional music along with the performance. The motion of the performer was also quantified by using background subtraction (Morita, Yamazawa, Terazawa, Yokoya, 2005). Moreover, particular musical pitches were assigned to each equal sized compartment on a screen. The notes implemented in a computer were played when the area-by-area amounts of movements exceed the criteria.

## 3. Method

### 3.1 Performer

The performer, Miyaji Katsura, was a male aged 35 years with a four-year career as a professional Rakugo artist. He had received the 27th NHK new entertainer award in 2012 and was one of the most popular young performers of the day.

### 3.2 Story and conventional rule of performing of Rakugo

In the experiment, he performed the *Rakugo* story called "*Omitate*" ("Your Choice" in English, Morioka & Sasaki, 1990). The story originally appeared in a book around the early 19th century, and was passed down orally. In the story, a young hired hand gets annoyed and plays on both sides of the fence between an egocentric rich man and a high-handed *Oiran* (a female entertainer). The performer, who took the atmosphere of the audience and the experimental settings into consideration, selected this particular story based on his intuitions. In the story, the performer has to play several characters, since the story is performed by just one performer. According to conventions of the Japanese theater, a person of high rank has to appear at the right to the center of the stage, as seen from the audience (Morisaki & Sasaki, 1990). To represent positional relationships when presenting a dialogue between people of different ranks, the performers have to move their head ceaselessly from left-high to right-low, and vice versa. This technique is known as "*Kamishimo*," and it describes how the performer switches between characters.

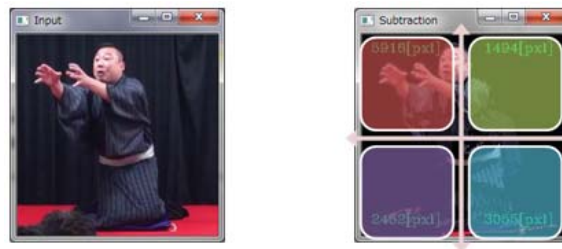
### 3.3 Participants in the Audience

Audience members were volunteer participants ( $N = 20$ , 10 men and 10 women, aged 16 to 67,  $M$  age = 40.6,  $SD = 16.4$ ), who were all familiar with *Rakugo* performances. The ratio of members that were familiar with the performer and the story were 35 % and 70 %, respectively.

### 3.4 Apparatus and procedure

The experiment took place in a laboratory (4.8 meters by 4.8 meters) with a purpose-build stage for *Rakugo* performance (1.8 meter width, by 1.8 meter length, by 0.9 meter height). At the rehearsal, the director asked the performer to play an actual part of a *Rakugo* performance, identical to a vaudeville setting. All the participants were invited to enter the room and sit on previously assigned chairs according to the order that participants applied to participate in the experiment. Then, participants were briefly familiarized with the room before the experiment was initiated. The director informed the participants that they would be record during the performance. The experimenters started to record and played the entrance music for the performer. Following this, the performer entered the stage and after he was seated, he made a deep bow and began a prologue called a *Makura*, meaning "pillow," which forms the basis for the main tale. *Makura* has the function of warming up the audience and leading them to the fictional world of the main plot (Morioka & Sasaki, 1990, pp. 26-27). Subsequently, the performer started telling the story "*Omitate*" which lasted about 30 minutes. This was all done according to the usual formalities and timing of legitimate *Rakugo* performances. The experimenters recorded sounds and images, as well as audience reactions during the performance. After the performance, the director asked the participants to inform the extent to which they enjoyed and appreciated the performance using a continuous scale ranging

from 0 (*no enjoyment*) to 100 (*full enjoyment*). Finally, the audiences completed a questionnaire that inquired about their familiarity with the performer, and the performance, prior to participating in the experiment.



**Figure 1.** An example of Input and Subtracted image

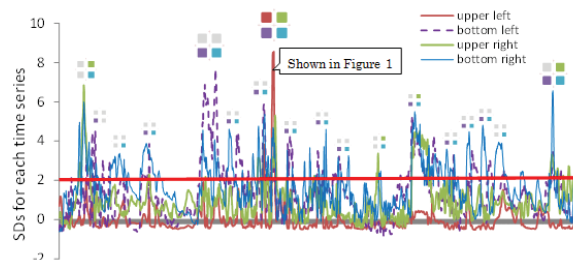
### 3.5 Coding and implementing

Each audience member was arbitrary assigned to one musical pitch on the keyboard prepared on the MIDI sequencer software (*Sekaiju*, OpenMIDIProject). The experimenter manually coded the eyeblinks of the audience by using ELAN 4.5.1 (Max Planck Institute for Psycholinguistics, Wundtlaan), which was developed for discourse analysis (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). When a member of the audience blinked, the assigned note corresponding to that member was commanded to play. Moreover, when many audience members blinked simultaneously, we could hear the sound of the blinks as a chord. Similarly, if an audience member blinked with a slight time lag, it sounded as a tuplet. Finally, the music that was composed was converted from MIDI to a wmv file for playback using the Windows Media Player.

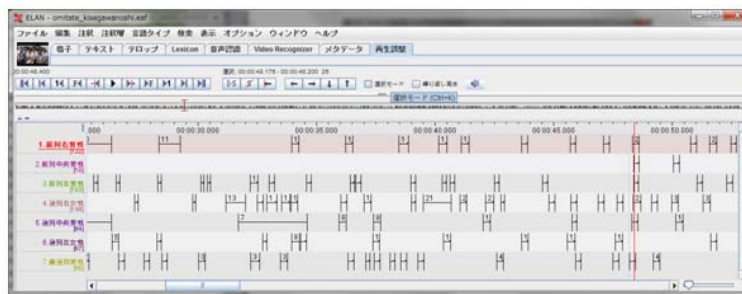
The motion of the performer was also quantified by using a background differencing technique, which is a popular technique in computer vision processing (Morita et al., 2005, see Figure 1). The screen was divided into equally-sized nine compartments and musical pitches were assigned to each area according to its location. The upper area was assigned to a higher pitch and the lower area to a lower pitch. These sounds had a different timbre from that of the audience. That is, we were able to distinguish between the sounds of the audience and the performer, even if they were identical in terms of intensity and pitch. The sounds for the performer made when the extent of area-by-area movements exceeded 2 standard deviations, indicative of a change that was much larger than usual body swaying. This calculation was performed by using a program developed by the author, based on Visual C++ 2012 and OpenCV2.4.5. The eyeblink music that was composed was then played with a subtracted movie of the performance. Following the quantification, the author manually checked the movie to identify the dominant modularity of scenes where the audience blinked collectively.

## 4. Results and Discussion

The mean score of enjoyment rating was very high ( $M = 82.0$ ,  $SD = 18.09$ , on a scale of 0 to 100). This suggests that most of the audience highly enjoyed the performance. Each responses of the audience often, or nearly coincided with little time difference. In the current experiment, eyeblinks increased when there were change to the scene, a switch of characters (i.e., *Kamishimo*), and shortly after the punch line that accompany humorous gestures and facial expression.



**Figure 2.** SDs of change-detected pixels of each area



**Figure 3.** An example of coded eyeblinks and their synchronization

The performance and the responses of the audience showed interchanges with each other, similar to leading and following, rather than a perfect on-time synchrony. As the eyeblinks of the audience could be related to their process of allocating attention, these results suggest that the performer loosely controlled whether or not the audience allocated their attention. In other words, a good performance literally "catches the eyes of audience". The dominant expressive modularity of the performer was movements of both hands. It made the performer appear as a conductor that orchestrated the eyeblinks of the audience. In an extreme instance, we could listen to the music played in accordance with interactions that have not been researched, such as when the performer just moves without saying a word.

Over the years, methods of capturing interactions between an audience and the performer have been a fundamental problem in research on the performing arts. The method proposed in the current study demonstrated that people could easily listen to an interaction as a temporal gestalt (i.e., music). This is owing to both the nature of sound and the innate and elaborate ear system. The method was useful, not because I dealt with a Japanese traditional *Rakugo* performance, but because eyeblinks inherently relate to the allocation of attention during the appreciation of all kinds of performing arts. This clearly suggests the possible application of the proposed method for quantifying and recognizing interactions in other types of performing arts. It is suggested that this method could help provide factual data on the expertise of performers and their key skills and arts.

## References

- Corey, D. P., García-Añoveros, J., Holt, J. R., Kwan, K. Y., Lin, S. Y., Vollrath, M. A., ... & Zhang, D. S. (2004). TRPA1 is a candidate for the mechanosensitive transduction channel of vertebrate hair cells. *Nature*, 432 (7018), 723-730.
- Hall, A. (1945). The origin and purposes of blinking. *The British journal of ophthalmology*, 29 (9), 445.
- Nakano, T., Kato, N., & Kitazawa, S. (2011). Lack of eyeblink entrainments in autism spectrum disorders. *Neuropsychologia*, 49 (9), 2784-2790.
- Nakano, T., Kato, M., Morito, Y., Itoi, S., & Kitazawa, S. (2013). Blink-related momentary activation of the default mode network while viewing videos. *Proceedings of the National Academy of Sciences*, 110 (2), 702-706.
- Nakano, T., & Kitazawa, S. (2010). Eyeblink entrainment at breakpoints of speech. *Experimental brain research*, 205 (4), 577-581.
- Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., & Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proceedings of the Royal Society B: Biological Sciences*, 276(1673), 3635-3644.
- Nomura, R., & Maruno, S. (2011). Constructing a Coactivation Model for Explaining Humor Elicitation. *Psychology*, 2 (5), 477-485.
- Morioka, H., & Sasaki, M. (1990). *Rakugo: the popular narrative art of Japan* (part 3). Cambridge: Council on East Asian Studies, Harvard University.
- Morita, S., Yamazawa, K., Terazawa, M., & Yokoya, N. (2005). Networked remote surveillance system using omnidirectional image sensor. *The Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, 88 (5), 864-875.
- Stern, J. A., Walrath, L. C. & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology*, 21, 22-33.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.